General Assembly

STUDY GUIDE

Legal frameworks for AI Governance

Address liability and accountability in AI-related accidents or decisions

St PETER'S SCHOOL
Barcelona

SMUN 2030
Singularity Model United Nations

TABLE OF CONTENTS

1. Letter from the chairs

Dear Delegates,

Hello everyone! We hope this letter finds you well as we prepare for the Singularity MUN 2024 conference. We are Andrea Pareja López and Manuela Modolin, your Chairs for the United Nations General Assembly (UNGA) committee, and we couldn't be more excited to embark on this journey with you.

Let us share a bit about ourselves. I'm Andrea, a 21-year-old final-year student pursuing a degree in Global Studies at Universitat Pompeu Fabra. I recently returned from a fantastic exchange semester in Geneva, Switzerland, which has given me a broader perspective on global issues. I've been passionate about Model United Nations since 2020, which led me to join the United Nations Student Association from my university.

I'm Manuela, I'm 17 years old and I study at St Peter's School Barcelona, where I am completing the IB course. I've participated in two MUNs as a delegate before, one in St Peters and one in New York where we got to visit the UN headquarters, which motivated me to participate as a co-chair this year. Additionally, as I want to study law, I believe this will help me improve my communication skills.

Our goal in creating this study guide is to assist you in navigating the intricate theme we'll be exploring: "Legal frameworks for AI Governance: Addressing liability and accountability in AI-related accidents or decisions." The evolving landscape of AI governance demands our careful consideration, and we're confident that together, we can delve into this topic with depth and innovation.

We encourage each of you to embrace collaboration fully. Let's seek common ground, build alliances, and work together to develop comprehensive and effective legal frameworks for AI governance.

If you have any questions or need clarification on committee matters, please feel free to reach out to us at andreaaparejaa@gmail.com or manumodolin07@gmail.com. We're here to support you on this MUN journey.

We're genuinely looking forward to your insightful contributions during our committee sessions. The Singularity MUN 2024 is bound to be a rewarding experience, and together, we're confident that we'll craft interesting solutions and resolutions.

Best regards,

Andrea Pareja López and Manuela Modolin
Chairs, Singularity MUN 2024 - UNGA Committee.

2. Committee competencies

Established in 1947, the United Nations General Assembly (UNGA) Sixth Committee, commonly known as the Legal Committee, plays a pivotal role in global legal affairs.

Composed of representatives from all 193 member states of the UN, UNGA operates in annual sessions to efficiently manage the vast array of global issues it addresses. This approach enables the UNGA to tackle current global challenges methodically, with each session focusing on specific agendas, led by elected officials such as the President and Vice Presidents. The session-based organization ensures continuity, progression in international diplomacy, and the ability to respond to evolving global dynamics. This means that the committee's leadership rotates each year.

Formal sessions are conducted with a high degree of protocol and order. Speeches and statements are usually delivered in one of the UN's six official languages: Arabic, Chinese, English, French, Russian, or Spanish. These statements are translated by UN translators to facilitate understanding among representatives from different countries. The order of speaking is typically arranged beforehand, giving priority to high-ranking officials from member states. This structured approach ensures that all nations have an opportunity to voice their perspectives in an organized and respectful environment.

President Dennis Francis from Trinidad and Tobago leads the current 78th session of the United Nations General Assembly, which began on September 5, 2023. This session features a diverse leadership with vice presidents from the five permanent members of the UN Security Council (China, France, Russia, the United Kingdom, and the United States) and 15 other nations. Similarly, observer states and supranational bodies play a significant role. Observer states, while not full members, are recognized entities that can participate in UNGA sessions but do not have voting rights (Holy See and Palestine). Supranational bodies are organizations formed by treaties between three or more nations that hold a certain level of authority over their member states. Technical and legal support is provided by the UN Secretariat's Legal Affairs Division, and member states often include legal experts in their delegations to address specific issues.

Tasked with supporting the General Assembly in enacting Article 13 of the United Nations Charter, the Legal Committee focuses on creating and formalizing international law. This includes the preparation of legal instruments and treaties. A notable early achievement of the committee was the drafting of the Convention on the Prevention and Punishment of the Crime of Genocide in 1948.

Additionally, the committee reviews reports from various entities, such as the United Nations Commission on International Trade Law (UNCITRAL), the Special Committee on the Charter of the United Nations and the Strengthening of the Organization, and the International Law Commission. The committee also contributes to advancing international law by discussing topics like the criminal accountability of UN mission officials, the rule of law, and the extent of universal jurisdiction, among others.

However, while the Legal Committee, like other General Assembly bodies (except for the Security Council), can create guidelines and regulations, it does not have the authority to enforce them. Instead, it depends on the willingness of individual countries to adopt and implement these guidelines and regulations. Despite this limitation, the Legal Committee, along with the other General Assembly bodies, continues to be of significant influence in international relations and law.

3. Definition of key terms

Accountability → refers to the obligation, responsibility, or answerability of individuals, organizations, or entities for their actions, decisions, or performance in a particular role or context. It involves being transparent about one's actions, accepting the consequences of those actions, and being willing to provide justifications or explanations when necessary. Accountability plays a crucial role in various fields, including governance, business, ethics, and law, as it ensures that individuals or entities are held responsible for their behavior and can be subject to appropriate consequences or measures in cases of wrongdoing or failure to meet expected standards.

AI Governance Framework → a structured set of regulations, policies, standards, and best practices. It governs AI technology development, application, and usage, ensuring that AI

systems are developed and used ethically, responsibly, and in compliance with legal standards.

AI regulation → involves the development of public sector policies and laws aimed at promoting and regulating AI, tying closely to the broader regulation of algorithms.

Algorithm → a step-by-step, systematic set of well-defined instructions or rules that are designed to solve a specific problem or perform a particular task. Algorithms serve as a fundamental concept in computer science and mathematics, guiding the process of transforming input data into desired output, often intending to achieve efficiency and accuracy. These instructions can encompass a wide range of actions, from basic arithmetic calculations to complex operations within computer programs. Algorithms are utilized across various domains, including data analysis, artificial intelligence, cryptography, and problem-solving, and they play a crucial role in the development of software and technology systems. The effectiveness and efficiency of algorithms are evaluated based on factors such as correctness, resource usage (time and space complexity), and scalability.

Cognitive Behavioral Manipulation → refers to a deliberate and often unethical practice that utilizes artificial intelligence (AI) and data-driven algorithms to influence and control individuals' thoughts, emotions, beliefs, and behaviors. E.g. Imagine a social media platform using cognitive behavioral manipulation to show tailored content that amplifies users' existing beliefs and preferences, reinforcing their biases and potentially leading to increased polarization and divisiveness within society.

Deepfake → a type of synthetic media, often in the form of audio, video, or images, created using advanced artificial intelligence techniques, particularly deep learning algorithms. Deepfakes involve the manipulation or fabrication of content in a manner that convincingly replaces, alters, or superimposes existing content with new, digitally generated material. These sophisticated AI-driven tools can convincingly replicate the appearance and behavior of individuals, including their facial expressions, voices, and gestures, making it challenging to discern the authenticity of the generated content. Deepfakes have raised concerns due to their potential for misuse, including the spread of disinformation, impersonation, and privacy violations, and have prompted increased scrutiny and efforts to develop detection and mitigation methods.

Deep Learning Algorithms → a subset of machine learning techniques that model and simulate the human brain's neural networks to enable artificial intelligence systems to learn

and make decisions from vast and complex data sets. These algorithms consist of multiple layers of interconnected artificial neurons, referred to as neural networks, which autonomously extract hierarchical and abstract features from data. Deep learning excels in tasks such as image and speech recognition, natural language processing, and pattern analysis. Deep learning algorithms have achieved remarkable success in various fields, including computer vision, autonomous driving, and medical diagnosis, thanks to their ability to automatically discover patterns and representations within data, leading to improved accuracy and performance in a wide range of applications.

Hard Law → refers to legally binding rules and regulations established through formal and recognized legal processes. These laws are enforceable by legal authorities and often include statutes, treaties, and formal regulations. Violations of hard law can lead to legal penalties or sanctions. Hard law is characterized by its definitive, clear obligations, and requirements, which are formally adopted and ratified by relevant legislative or regulatory bodies.

Liability → a legal concept that denotes the legal responsibility or obligation of an individual, organization, or entity to compensate or bear the consequences of harm, losses, or damages caused to another party due to their actions, negligence, or legal obligations. It encompasses the financial or legal responsibility for injuries, losses, or breaches of duty and can result in monetary compensation or other legal remedies to redress the harm inflicted.

Machine Learning → a subfield of AI that focuses on the development of algorithms and statistical models that enable computer systems to automatically learn and improve from experience or data without being explicitly programmed. In essence, machine learning empowers machines to recognize patterns, make predictions, and make decisions based on past examples or information. It involves the use of data to train and refine models, allowing computers to generalize from that data and make informed decisions or predictions when faced with new, unseen data. Machine learning has a wide range of applications, including natural language processing, image recognition, recommendation systems, autonomous vehicles, and predictive analytics, and it continues to play a vital role in advancing AI technologies and solving complex problems in various domains.

Negligence → a legal concept that refers to the failure to exercise reasonable care or attention, resulting in harm or damage to another person or their property. It involves a breach of the duty of care owed to others, where individuals or entities do not take the precautions or actions expected of them, leading to potentially avoidable harm or losses.

Negligence is a fundamental principle in tort law and can give rise to legal liability when it is proven that a duty of care was breached, causing harm to another party.

Social Scoring → a data-driven system that evaluates and assigns numerical scores or ratings to individuals based on their online and offline behaviors, activities, and interactions within a society or community. E.g. In some regions, governments or private companies implement social scoring systems that assess citizens' behavior, including online activities, financial transactions, and civic engagement. Individuals with high social scores may receive benefits like faster access to loans or preferred job opportunities, while those with low scores may face restrictions on travel or access to certain services.

Soft Law → consists of guidelines, principles, or standards that, while not legally binding, are influential in shaping behavior and practices. Soft law instruments, such as policy declarations, codes of conduct, or best practice guidelines, are adopted through less formal processes. Although they lack enforceable legal authorities, they carry moral or political weight, influencing and guiding behavior in various fields. Compliance with soft law is voluntary, and it is often used in areas where formal, binding legislation is difficult to establish or enforce.

Tort Law → a branch of civil law that addresses situations where a person's actions or inactions cause harm, loss, or injury to another individual. It focuses on providing a remedy, usually in the form of monetary compensation, to the injured party. Tort law is based on making the injured party 'whole' again, to the extent possible, through financial reparation. It covers a wide range of wrongful acts, including negligence, intentional harm, and strict liability offenses, where the defendant can be held responsible regardless of intent or negligence. The key components of a tort typically involve proving that there was a duty of care, a breach of that duty, causation linking the breach to the harm, and actual damage or harm suffered by the plaintiff. Tort law serves as a mechanism to enforce personal responsibility and social norms, and to deter harmful behavior.

4. History of the topic

a. History of AI

The history of AI is marked by significant milestones that have reshaped both consumer lifestyles and business operations. From the introduction of the Turing Test in 1950 to the celebrated launch of ChatGPT, AI has grown into a powerful force in our daily lives and the global economy.

AI's journey began in the 1950s, paralleling the dawn of the Atomic Age. Both AI and atomic technology have been viewed as potential existential threats, but AI's advancements have been remarkable despite intermittent funding challenges. Its impact became more visible to the public over the past decade, with products like Apple's Siri and Amazon's Alexa, online shopping, social media, and self-driving cars altering everyday life and business practices.

The evolution of AI has included several key breakthroughs. In the 1950s, the terms "artificial intelligence" and "machine learning" were coined, and the first artificial neural network was developed. The 1960s saw the creation of Eliza, a chatbot with cognitive abilities, and Shakey, the first mobile intelligent robot. The 1970s and 1980s experienced AI winters followed by renaissance. The 1990s brought advances in speech and video processing. The 2000s were marked by the rise of IBM Watson, personal assistants, facial recognition, deepfakes, autonomous vehicles, and advancements in content and image creation.

Today, AI continues to evolve rapidly, with organizations increasingly considering AI in their operational strategies. Gartner predicts that by 2026, AI models focused on transparency, trust, and security will significantly improve adoption and user acceptance. As AI progresses towards achieving artificial general intelligence, neuromorphic processing, and other advancements raise critical issues around trust, privacy, transparency, accountability, ethics, and humanity. These issues will continue to be central as AI shapes various sectors, including business, healthcare, education, agriculture, law, and transportation.

b. History of AI regulation and legal frameworks

The global regulatory landscape for AI is an emerging concern across various jurisdictions, including the European Union, which possesses governmental regulatory authority, and through supra-national bodies like IEEE and OECD, as well as intergovernmental organizations like the United Nations.

AI regulation focuses on the risks and biases of machine-learning algorithms, emphasizing the need for transparency, explainability, and accountability. Proposals range from hard law approaches, which face challenges due to the rapid evolution of AI, to soft law approaches that offer greater adaptability but lack enforcement power. An alternative model suggests using intellectual property rights to enforce ethical practices in AI.

Since 2016, there has been a significant rise in AI ethics guidelines, aimed at maintaining social control over AI technology. Regulation is deemed necessary to both foster AI development and manage its associated risks. Moreover, organizations deploying AI are expected to play a pivotal role in creating and deploying trustworthy AI, adhering to principles of ethical AI, and taking responsibility to mitigate risks. The concept of AI regulation through mechanisms such as review boards is viewed as a societal approach to addressing the AI control problem.

As a result, there has been a notable increase in AI-related laws worldwide. For instance, the AI Index at Stanford reported a leap from one AI-related law passed in 2016 to 37 in 2022 in the surveyed countries. The need for AI regulation has been recognized since the 1960s, with advocates in responsible AI, AI ethics, consumer protection, and cybersecurity vocalizing the need for regulatory guardrails around AI development. High-profile figures like Elon Musk have called for AI regulation, underscoring the risks of unregulated AI. This call for regulation has been met with mixed responses, ranging from skepticism about regulating a technology still in development to suggestions for developing common norms and requirements for AI systems.

Public attitudes towards AI vary significantly across countries, with varying degrees of perceived benefits and risks. For instance, a significant portion of the American public recognizes AI as posing risks to humanity, with many supporting the need for federal regulation. Different countries have approached AI regulation in varied ways, reflecting their economic and political landscapes. The US, China, and the EU each have distinct approaches, ranging from market-driven to state-driven and rights-driven strategies. Nations like Australia and Brazil are formulating their AI strategies, with Brazil focusing on ethical principles and Australia calling for a national AI task force.

Initiatives like the Global Partnership on Artificial Intelligence and the International Telecommunication Union's AI for Good platform represent global efforts to align AI development with human rights and democratic values. These initiatives aim to harness AI

for global development goals and to address challenges like the COVID-19 pandemic. Similarly, the regulation of lethal autonomous weapons systems (LAWS) has been under discussion at the United Nations since 2013, with concerns about compliance with laws of armed conflict and the adequacy of existing international law to address the development of fully autonomous weapons.

The history of AI regulation and legal frameworks is marked by a growing awareness of the need for governance in this field. As AI continues to advance and permeate various sectors, the development of comprehensive, adaptable, and ethical regulatory frameworks will be essential for harnessing the benefits of AI while mitigating its risks. This evolving regulatory landscape presents both challenges and opportunities for policymakers, industries, and society as a whole.

c. History of AI Accountability and Liability

The intersection of AI and law has become increasingly significant, particularly in the context of AI-driven accidents. As AI technologies evolve and integrate into various aspects of life, their potential to cause environmental, societal, and personal harm has become increasingly debated.

AI-driven accidents, such as those involving self-driving cars, have highlighted the complexities in applying traditional legal norms due to AI's opaque decision-making processes and the distributed responsibility among different actors involved in AI development and deployment. The year 2021 saw almost 400 car crashes involving self-driving cars in the United States alone, underscoring the urgency of legal adaptation to AI technologies.

Tort law, which deals with acts or omissions causing harm or injury, plays a crucial role in addressing AI accidents. However, AI systems challenge the traditional mechanisms of tort law, particularly the concept of negligence, which requires proving causation. AI's inscrutable decision-making and the diffusion of responsibility among multiple actors make establishing causation and foreseeability challenging.

AI systems involve multiple actors, including hardware manufacturers, software developers, and data trainers, leading to the "problem of many hands" – a diffusion of responsibility. This complexity raises questions about who should be held liable in AI accidents. In 2017, the

European Parliament considered a compensation fund financed by those involved in AI systems to compensate victims, offering a balance between guaranteed compensation and determining responsibility. The 2017 proposal of granting "electronic personality" to robots faced significant backlash and was ultimately dropped, alongside the idea of a compensation fund. Such concepts highlighted the need for a careful balance between incentivizing safety in AI products and ensuring accountability.

As such, the accountability landscape in AI involves various stakeholders, each with distinct roles and responsibilities:

1. AI users: They bear primary responsibility for the proper operation of AI systems, and understanding their functionality and limitations.
2. AI users' managers: These individuals ensure their teams are trained to use AI responsibly and adhere to company AI policies.
3. AI users' companies/employers: Organizations using AI must establish guidelines for its application and are accountable for AI-related outcomes within their operations.
4. AI developers: This group, including entities like OpenAI, is accountable for developing AI responsibly, ensuring unbiased design, and incorporating safety measures.
5. AI vendors: Providers of AI products or services must offer reliable and ethical solutions and be accountable for product flaws or undisclosed risks.
6. Data providers: These entities are responsible for the quality and ethical sourcing of data used by AI systems.
7. Regulatory bodies: They are tasked with creating and enforcing laws governing AI use, thus playing a crucial role in defining the broader legal and ethical landscape for AI accountability. Delegates are part of this stakeholder.

5. Timeline of international action

- **2016:** European Parliament Resolution - The European Parliament passed a resolution on Civil Law Rules on Robotics, calling for EU-wide rules and asking the European Commission to consider a specific legal status for robots.
- **2016:** IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems - IEEE has developed ethically aligned design guidelines for AI and autonomous systems, focusing on ethical values, accountability, and transparency.

- **2017:** Asilomar AI Principles - Developed by the Future of Life Institute, these principles are a set of guidelines for AI research and development, including aspects of liability and ethics.

- **2018:** Montreal Declaration for Responsible AI - This declaration emphasizes the ethical development of AI and includes principles related to responsibility and the impact of AI on society.

- **2018:** EU General Data Protection Regulation (GDPR) - The GDPR regulates data protection and privacy in the European Union and the European Economic Area. It addresses the processing of personal data within the EU, including data used by AI systems, ensuring transparency, security, and accountability.

- **2019:** OECD Principles on Artificial Intelligence - These principles, adopted by OECD member countries, provide guidelines for responsible stewardship of trustworthy AI, addressing aspects of accountability and transparency.

- **2019:** EU Ethics Guidelines for Trustworthy AI - Released by the High-Level Expert Group on AI appointed by the European Commission, these guidelines include recommendations for ensuring accountability in AI systems.

- **2020:** The Rome Call for AI Ethics - Endorsed by the Vatican and major tech companies, this call emphasizes ethics in AI, including fairness, transparency, and responsibility.

- **2021:** AI and Human Rights in Trade Policy - This joint declaration by multiple countries at the WTO Public Forum discussed the impact of AI on trade, emphasizing accountability and transparency.

- **2021:** Proposal for AI Regulation by the European Commission - This proposed regulation is one of the first major legal frameworks specifically addressing AI, focusing on high-risk AI systems and including requirements for transparency, accountability, and human oversight.

- **2022:** UNESCO Recommendation on the Ethics of AI - This document provides a global standard for the regulation of AI, focusing on ethical principles, human rights, and values.

- **2022:** EU AI Liability Directive -This directive introduces a rebuttable presumption of causality in AI-related accidents, linking the breach of a duty of care with the AI system's output. This presumption is designed to ease the burden on victims in proving causation in AI-related incidents.

- **2022:** EU Product Liability Directive - This directive proposes amendments to include AI products in the product liability regime. It enables courts to establish causal links between defectiveness in AI products, causality, and injury, simplifying the process for victims to claim compensation.

- **2022:** Canada's Artificial Intelligence and Data Act (AIDA) - Proposed as part of Canada's Digital Charter, AIDA aims to ensure AI systems are used responsibly, addressing accountability, transparency, and the protection of personal information.
- **2023:** EU Artificial Intelligence Act - Provisional agreement between the Council presidency and the European Parliament, aiming to ensure AI systems in the EU market are safe and respect fundamental rights and EU values.

6. Timeline of UN action

- **November 2017:** The United Nations System Chief Executives Board for Coordination (CEB) begins examining risks and opportunities associated with new and emerging technologies, including AI. This marks the initiation of the UN's focused efforts on AI's impact on sustainable development.
- **2017-2019:** The High-level Committee on Programmes (HLCP) undertakes analyses and develops strategies on frontier technologies, including AI. These analyses are intended to guide the UN's approach to these rapidly evolving technologies.
- **May 2019:** The HLCP completes work on key frontier topics like artificial intelligence, the future of work, and innovative education, presenting system-wide strategies for endorsement by the CEB.
- **April 2019 (37th Session of HLCP):** The HLCP members approve the United Nations system-wide strategic approach and road map for supporting capacity development on artificial intelligence. This strategy emphasizes AI-related capacity-building for developing countries, stakeholder engagement, and ethical development of AI.
- **2019 (CEB's First Regular Session):** CEB endorses the system-wide strategic approach and road map for AI capacity development, underscoring the commitment of the UN system to promote AI as a force for good.
- **November 2019:** UNESCO General Conference mandates the organization to support the elaboration of a non-binding Recommendation on the ethics of AI.
- **May 2020:** The first version of UNESCO's Recommendation on the ethics of AI is published, and prepared by an Ad Hoc Expert Group.
- **July 2020:** HLCP approves the UN system-wide contribution to UNESCO's Ad Hoc Expert Group for the preparation of the Recommendation on the ethics of AI.

- **October 2020 (40th Session of HLCP):** Creation of the Inter-agency Working Group on AI (IAWG-AI), co-led by UNESCO and ITU, to consolidate UN expertise on AI and support CEB and HLCP workstreams on AI ethics and capacity development.
- **November 2021:** UNESCO's General Conference adopts the Recommendation on the Ethics of AI, a significant milestone in establishing ethical guidelines for AI.
- **July 2022:** HLCP approves the Principles for the Ethical Use of Artificial Intelligence in the United Nations System, translating UNESCO's Recommendation into actionable principles for UN entities.
- **September 2022:** The CEB endorses the Principles for the Ethical Use of AI in the UN System.

7. Current situation

In recent years, artificial intelligence (AI) has risen as a transformative force in the technological landscape, reshaping industries and altering how we interact with the world. AI's capabilities in automating procedures and revolutionizing connections demonstrate its immense potential. However, this rapidly advancing technology brings forth significant ethical, legal, societal, and organizational challenges. As AI systems become deeply integrated into daily life, the urgency for a comprehensive AI governance framework grows stronger.

The absence of effective AI governance poses risks such as privacy violations, biased algorithms, and the potential misuse of AI for harmful purposes. A robust AI governance framework is essential to ensure transparency, accountability, and responsible development and deployment of AI technologies. This framework is crucial for mitigating risks and harnessing AI's potential in enhancing productivity and innovation. According to McKinsey & Company, Generative AI alone could significantly boost business revenues, highlighting the importance of balancing AI advancement with governance.

Components of an effective AI Governance framework
A comprehensive AI governance framework includes several key components:

1. *Establishing ethical guidelines:* Defining universal principles and values for AI systems, including fairness, transparency, accountability, and privacy.
2. *Ensuring data security:* Safeguarding data privacy and security, and ensuring consent for data collection and sharing.
3. *Maintaining transparency:* Being open about the AI model's purposes, data collection, and processing.
4. *Demonstrating accountability:* Setting clear guidelines for liability associated with AI developers and systems.
5. *Mitigating discrimination:* Developing strategies to identify and reduce biases to prevent unfair outcomes.
6. *Regulation and compliance:* Adhering to data protection and AI laws to avoid penalties.
7. *Monitoring and assessment:* Continuously monitoring AI systems and conducting risk assessments to address vulnerabilities.

Building an effective framework involves assessing organizational needs, managing sensitive data responsibly, understanding legal standards, ensuring transparency in AI decision-making, and establishing clear roles and accountability. It also includes developing data management policies, documenting processes, and adapting to the rapidly evolving AI field. Unregulated AI growth can lead to massive-scale unauthorized surveillance, data breaches, cultural biases in critical scenarios, and detailed behavioral profiling. The risks have become so pronounced that in March 2023, thousands of technology experts called for a halt in AI development unless properly regulated.

To adapt tort law to AI's characteristics, the European Union proposed the AI Liability Directive and the Product Liability Directive in 2022. These directives aim to ease the burden on victims to establish causation in AI accidents. Notable proposals include:

- Rebuttable presumption of causality: This presumption links the breach of a duty of care with the AI system's output.
- Facilitated evidence access: The directives propose granting courts the power to order disclosure of evidence related to high-risk AI systems, enhancing transparency and enabling a clearer understanding of AI decisions.
- Product liability amendments: Proposed amendments to include AI products in product liability regimes would allow courts to establish causal links between defectiveness, causality, and injury.

As a result, regulations are vital for ensuring ethical AI deployment, establishing safety and accountability standards, protecting data privacy, addressing biases, and fostering transparency, while simultaneously supporting innovation, international collaboration, consumer trust, and cybersecurity, and providing legal frameworks for AI-related incidents.

EU Artificial Intelligence Act

This recent provisional agreement between the Council presidency and the European Parliament marks a pivotal advancement in AI regulation. Its main objective is to ensure AI systems in the EU market are safe and respect fundamental rights and EU values. This is extremely significant as it is the first legislative proposal of its kind globally, potentially setting a global standard for AI regulation.

Key elements of the Agreement

1. Risk-based approach: Classifying AI systems based on their potential to harm society, with stricter rules for higher-risk AI.

2. Definitions and scope: Aligning AI system definitions with the OECD approach, excluding military/defense purposes, and focusing on research, innovation, and non-professional use.

3. High-risk and prohibited AI practices: Establishing a high-risk classification with obligations for market access and banning certain AI practices like cognitive manipulation and untargeted scraping of facial images.

4. Law enforcement exceptions: Allowing emergency use of high-risk AI tools by law enforcement, with specific mechanisms to protect fundamental rights.

5. General-purpose AI systems and foundation models: Introducing specific rules and transparency obligations for general-purpose and foundation AI models, particularly for those with high impact.

6. New governance architecture: Setting up an AI Office within the Commission for overseeing advanced AI models, with support from a scientific panel and an AI Board for member state coordination.

7. Penalties: Establishing fines based on global annual turnover or a fixed amount, with proportionate caps for SMEs (small and medium-sized enterprises) and start-ups.

8. Transparency and fundamental rights: Requiring a fundamental rights impact assessment and increased transparency for high-risk AI systems.

9. Innovation support: Modifying provisions to create an AI regulatory sandbox and allowing real-world testing under certain conditions.

One of the highlights of this act is the introduction of a groundbreaking approach that distinguishes between different risk levels associated with artificial intelligence systems. This stratification of AI systems according to their perceived risk levels is set to have a profound impact on the legal framework surrounding AI, particularly in addressing issues of liability.

AI systems will undergo thorough evaluation and classification based on the risks they pose to users. These differentiated risk levels will dictate the degree of regulation applied, marking a significant departure from previous uniform approaches to AI governance. The Act is designed to safeguard the interests of the public while fostering innovation and responsible AI development.

At one end of the spectrum, AI systems categorized as "unacceptable risk" will face strict measures, with certain systems, such as those involving cognitive behavioral manipulation or social scoring, being completely banned due to their potential harm to individuals or

specific vulnerable groups. However, exceptions may be permitted for specific law enforcement purposes, underlining the careful balance between security and privacy concerns.

In contrast, AI systems categorized as "high risk" will be subject to a more comprehensive evaluation process. These systems, including those used in product safety, critical infrastructure management, and legal interpretation, will undergo scrutiny both before entering the market and throughout their lifecycle. This heightened oversight aims to ensure the safety and fundamental rights of users while maintaining the integrity of critical sectors.

For AI systems classified as "limited risk," minimal transparency requirements will be enforced, allowing users to make informed decisions about their interactions with AI applications, particularly those generating or manipulating content like deepfakes. This approach emphasizes user awareness and choice, promoting responsible usage without imposing unnecessary restrictions.

In general, the Act aims to balance fostering innovation and ensuring safety and compliance with fundamental rights. It provides a structured legal framework to promote AI investment and innovation across the single market.

8. Bloc positions

Bloc positions refer to the different "groups" countries could make based on their similar ideas on the topic, as different countries and regions have adopted varied approaches to AI regulation.

Innovation-driven bloc

Likely to advocate for less restrictive laws to promote AI development.

Israel:
Israel is recognized globally for its robust tech industry and significant advancements in AI and cybersecurity. It has a strong culture of innovation, particularly in the tech sector.

Japan:
Japan strongly emphasizes technological innovation, particularly in robotics and AI. The country has been a pioneer in robotics, and its strategies towards AI revolve around enhancing economic growth and societal welfare, placing it in the innovation-driven bloc.

South Korea:
South Korea has a strong focus on technological advancement, particularly in areas like AI, 5G, and robotics. The country invests heavily in research and development, aiming to be a leader in AI technologies.

USA:
Follows a market-driven approach (e.g. OpenAI and Microsoft). In the USA, discussions on AI regulation have covered a wide range of topics, from the nature of the regulatory framework to the roles of state governments and courts. The Obama administration set precedents for AI research and regulation, focusing on public safety and risk assessment. Various government agencies, including the Food and Drug Administration, have been involved in AI regulation, particularly in medical applications.

UK:
The UK has supported AI applications in business and public sectors, with strategies and guidance focused on responsible AI design and cyber security. The National AI Strategy outlines actions to assess AI risks, including catastrophic risks related to AGI.

Privacy-conscious bloc

May push for stringent regulations focusing on data protection and individual rights.

EU:

Follows a rights-driven approach. The Council of Europe aims to align AI with human rights and democratic values. The European Union plays a significant role in AI regulation, with the GDPR, Digital Services Act, Digital Markets Act, and the Artificial Intelligence Act, which is the most comprehensive AI regulation globally.

Singapore:

Known for its structured approach to governance, Singapore has been proactive in implementing AI while being mindful of privacy and ethical considerations. The city-state has a balanced approach, ensuring that its AI advancements go hand-in-hand with data protection and privacy laws.

Neutral or mediating countries

May seek balanced approaches and act as mediators between opposing blocs.

Canada:

Canada's Pan-Canadian Artificial Intelligence Strategy, initiated in 2017, focuses on increasing AI research and reflecting Canadian values in AI advancements. The strategy encompasses various programs and councils aimed at fostering responsible AI development.

Switzerland:

Switzerland is known for its neutrality in international affairs, and this extends to its stance on AI governance. It focuses on balancing innovation with ethical considerations. Switzerland's approach to AI is characterized by promoting research and development while also being attentive to privacy, ethical standards, and international collaboration. Switzerland's role in hosting international dialogues and its global reputation for diplomacy further reinforce its status as a neutral or mediating country in the realm of AI governance.

Developing nations

Could focus on equitable access to AI technology and mitigating the digital divide.

Brazil:

The Brazilian Chamber of Deputies approved the Brazilian Legal Framework for Artificial Intelligence in 2021, focusing on ethical AI development and addressing issues like non-discrimination and algorithmic accountability. However, this framework faced criticism for its lack of binding clauses and comprehensive coverage of accountability and transparency principles.

India:

India is emerging as a significant player in AI, primarily due to its large pool of IT professionals and rapid digital transformation. However, it still faces challenges typical of developing nations, such as infrastructure development and digital divide issues. India's focus is on leveraging AI for economic growth and addressing societal challenges.

Authoritarian regimes

Would emphasize state control and the use of AI for security, potentially clashing with privacy and human rights concerns.

China:

Follows a state-driven approach. China's AI regulation is driven by the State Council's "Next Generation Artificial Intelligence Development Plan," emphasizing state control over AI development and data. Ethical guidelines for AI use have been introduced, including control over data and adherence to national standards.

Russia:

The country has been investing significantly in AI technology, often with a focus on military and surveillance applications. The Russian government exercises strict control over AI development and deployment, aligning it closely with state interests and national security objectives. There is also a notable lack of transparency and limited public or international collaboration in AI initiatives.

9. Issues & questions to be tackled during sessions

Key issues for discussion:

- <u>Defining AI liability:</u> Establishing clear definitions and boundaries for liability in AI-related incidents.
- <u>Regulatory frameworks:</u> Examining existing regulations and their effectiveness in governing AI.
- <u>Ethical considerations:</u> Addressing ethical dilemmas posed by AI, including privacy, bias, and decision-making autonomy.
- <u>Global standards and cooperation:</u> Discussing the need for global standards and international cooperation in AI regulation.
- <u>Role of the private sector:</u> Analyzing the responsibilities and accountability of AI developers and corporations.
- <u>Transparency and explainability:</u> Ensuring AI systems are transparent and their decisions can be explained and understood.
- <u>Consumer protection and rights:</u> Focusing on the protection of consumer rights in the context of AI technologies.
- <u>Emerging technologies and future trends:</u> Anticipating future challenges and developments in AI technology.
- <u>Adapting traditional tort law to AI's unique challenges:</u> Ensuring justice and accountability for victims of AI accidents.

Questions a resolution must answer:

- How can international law be adapted to address the unique challenges posed by AI in terms of liability and accountability?
- What are the best practices for ensuring that AI developers and users are held accountable for AI-related accidents or decisions?
- How should liability be distributed among various stakeholders (developers, users, regulators) in the case of AI-related incidents?
- What ethical guidelines should govern the development and use of AI to protect individual rights and prevent discrimination?
- How can transparency and explainability in AI systems be improved to ensure accountability?

- How can consumer protection be balanced with innovation in the field of AI?
- What mechanisms can be implemented to monitor and regulate emerging AI technologies, such as deep learning and autonomous systems?
- Should there be a global consensus on AI ethics, and if so, what principles should it include?
- How can the private sector be effectively regulated to prevent abuse of AI technology while encouraging innovation?

10. Bibliography and further reading

AI Accountability: Who's Responsible When AI Goes Wrong?. (n.d.). Emerge Digital.
https://emerge.digital/resources/ai-accountability-whos-responsible-when-ai-goes-wrong/#:~:
text=AI%20Developers%3A%20AI%20accountability%20extends,to%20prevent%20misuse
%20or%20errors

*Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Worldwide Rules
for AI.* (2023). Council of the European Union.
https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-ac
t-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

Artificial Intelligence. (n.d.). United Nations System Chief Executives Board for Coordination
(UNSCEB). https://unsceb.org/topics/artificial-intelligence

AI Index Report 2023. (2023). Stanford Institute for Human-Centered Artificial Intelligence
(HAI).
https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

Chui et al. (2023). *The economic potential of generative AI: The next productivity frontier.*
McKinsey.
https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-o
f-generative-ai-the-next-productivity-frontier#key-insights

Delcker, J. (2018). *Europe divided over robot 'personhood'*. Politico.
https://www.politico.eu/article/europe-divided-over-robot-ai-artificial-intelligence-personhood/

EU AI Act: first regulation on artificial intelligence. (2023). European Parliament.
https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-fir
st-regulation-on-artificial-intelligence

European Commission. (2022). *Proposal for a directive of the european parliament and of
the council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability
Directive).*
https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf
European Commission. (2022). *Proposal for a directive of the european parliament and of
the council on liability for defective products*.

https://single-market-economy.ec.europa.eu/system/files/2022-09/COM_2022_495_1_EN_ACT_part1_v6.pdf

Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026. (2023). Gartner.
https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026#:~:text=By%202026%2C%20organizations%20that%20operationalize,for%20Generative%20AI%2C%202023.%E2%80%9D

Graziosi, G. (2022). *Self-driving car systems were involved in 400 crashes since 2021 - 273 were Teslas.* Independent.
https://www.independent.co.uk/news/world/americas/self-driving-car-crashes-tesla-b2101763.html

How to Develop an Effective AI Governance Framework?. (2023). Securiti.ai.
https://securiti.ai/ai-governance-framework/

Karjian, R. (2023). *The history of artificial intelligence: Complete AI timeline.* TechTarget.
https://www.techtarget.com/searchenterpriseai/tip/The-history-of-artificial-intelligence-Complete-AI-timeline

Mohammad et al. (2023). *An open letter signed by tech leaders, researchers proposes delaying AI development.* NPR.
https://www.npr.org/2023/03/29/1166891536/an-open-letter-signed-by-tech-leaders-researchers-proposes-delaying-ai-developme

Tong, A. (2023). *AI threatens humanity's future, 61% of Americans say: Reuters/Ipsos poll.* Reuters.
https://www.reuters.com/technology/ai-threatens-humanitys-future-61-americans-say-reutersipsos-2023-05-17/

Vasudevan, A. (2023). *Who is Liable for AI-Driven Accidents? The Law is Still Emerging.* Centre for International Governance Innovation (CIGI).
https://www.cigionline.org/articles/who-is-liable-for-ai-driven-accidents-the-law-is-still-emerging/